(2)

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

AGENCY USE ONLY	REPORT DATE	REPORT TYPE AND DATES COVERED
	30 Sept. 92	Final 1 June 89 - 31 Aug. 92

TITLE AND SUBTITLE

Effects on Learning of Individual Differences in Inferencing Ability

FUNDING NUMBERS

PE-61102F
PR 2313
TA-A9
AFOSR 89-0515

AUTHOR(S)

Bruce K. Britton

University of Georgia Research Foundation, Inc,.
University of Georgia
Athens, GA 30602-3013

AFOSR-TR- 93 0012

Air Force Office of Scientific Research
Bolling Air Force Base
Washington, DC 20332-6448
Dr John F. Tangney

DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release
Distribution Unlimited

93 2 01 061

In an earlier report, we accounted for about 83% of the variance in learning from instructional text, by regression equations using cognitive gaps in the text, with gap frequency measured by Kintsch's computer model of reading. The present studies tested three ways to improve Air Force recruits' learning from instructional text: (1) We developed an individual differences model of recruits' abilities needed to overcome the negative effects of the cognitive gaps in text. Tests on 211 Air Force recruits showed the model accounted for 54% of the variance in learning. The individual differences tests can be used to select and classify recruits; (2) We developed cognitive science methods of revising Air Force instructional texts to improve their learnability. The methods more than doubled the learnability of the revised text; and (3) We developed two computer job aids that editors can use to improve Air Force instructional texts by implementing our revision methods. Tests showed that (a) cognitive gaps were present at high frequency in each of a random sample of Air Force texts, and (b) measures calculated by the job aids were correlated as much as .62 with learning from instructional text. The revision methods can be used to improve most Air Force instructional texts.

SUBJECT TERMS

Learning, Individual Differences, Inference, Comprehension, Working Memory, Prior Knowledge

SECURITY CLASSIFICATION OF REPORT	SECURITY CLASSIFICATION OF THIS PAGE	SECURITY CLASSIFICATION OF ABSTRACT	LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unclassified

FO 1 SEP

Summary

In an earlier report, we accounted for about 83% of the variance in high-school graduates' learning from instructional text, by using the frequency of cognitive gaps in the text, with gap frequency measured by Kintsch's computer model of reading (Britton, Van Dusen, Glynn, & Hemphill, 1990; Miller & Kintsch, 1980). The present studies developed and tested three ways to use this fact to improve Air Force recruits' learning from instructional text.

1. We developed an individual differences model of recruits' abilities that are needed to overcome the negative effects of the cognitive gaps in text. Tests on 211 Air Force recruits showed that the model accounted for 54% of the variance in learning, and had a LISREL Goodness of Fit index of .97. The individual differences tests can be used to select and classify recruits to improve training efficiency and effectiveness.

2. We developed cognitive science methods of revising Air Force instructional texts to improve their learnability. The revision methods were based on Kintsch's computer model of reading. A test of the methods showed that they more than doubled the learnability of the revised text (tested on 213 Air Force recruits) over the original Air Force text (tested on 211 Air Force recruits).

3. We developed two computer job aids that editors can use to improve Air Force instructional texts by implementing our revision methods. Tests of these job aids showed that (a) cognitive gaps were present at high frequency in each of a random sample of Air Force texts, and (b) measures calculated by the job aids were correlated as much as .62 with learning from instructional text. This indicates that the revision methods can be used to improve most Air Force instructional texts.

Future development work is proposed to implement these techniques throughout the Air Force instructional text community, in order to produce large improvements in Air Force recruits' learning.

PREFACE

Development of this paper was supported by the Air Force Office of Scientific Research (AFOSR funding numbers PE-61102F, PR-2313, and TA-A9); and the research was conducted at the Air Force Learning Abilities Measurement Program (LAMP), a multi-year basic research program at the Air Force Human Resources Laboratory (AFHRL) which is sponsored by AFOSR. The goals of the LAMP program are to specify the basic parameters of learning ability, to develop techniques for the assessment of individuals' knowledge and skill levels, and to explore the feasibility of a model-based system of psychological measurement.

Support for the work documented here was provided by a contract from AFOSR to the University of Georgia Research Foundation, with Dr. Britton as principal investigator.

The author wishes to thank Terry Perdue and other members of the OAO Corporation for their assistance, and Patrick Kyllonen, Bill Tirre, Valerie Shute, Dan Woltz, Ray Christal, Abe Tesser, Sami Gulgoz, Lani Van Dusen, Barry Stennett, Cathy Brooks, and Pam Tidwell for their contributions. Correspondence concerning this article should be addressed to Bruce K. Britton, Department of Psychology, University of Georgia, Athens, GA 30602-3013.

AFOSR Technical Report 30 September 1992

Effects on Learning of Individual Differences in Inferencing Ability

Bruce K. Britton
Department of Psychology
University of Georgia
Athens, GA 30602-3013

TABLE OF CONTENTS

List of Figures

List of Tables

I. INTRODUCTION

The goal of this project was to develop broadly useful ways of improving Air Force recruits' learning from Air Force instructional text. We achieved this goal in three steps. First we identified the individual differences abilities that recruits need for learning. Second, we reduced recruits' needs for those abilities by targeted revisions of Air Force instructional texts. This doubled learning, especially among less able recruits. Third, we constructed computer job aids that make it possible for Air Force writers to improve texts as we did. Specifically:

A. We identified the individual differences abilities that determine recruits' learning from Air Force instructional text. When our individual differences model was tested on 211 Air Force recruits, it had a LISREL goodness of fit index of .97, indicating that it successfully specified the abilities and their relations. The tests for these abilities can be used to select and classify recruits, leading to a better fit between recruits' abilities and their training challenges.

B. We improved an Air Force instructional text by revising it so that recruits with lower levels of the individual differences abilities could learn much more easily from it. When our revised text was tested against the Original Air Force text on 213 Air Force recruits, the results showed that learning was more than doubled for the Revised Version over the Original Version. Doubling learning could halve the costs of training. Our techniques for revising can be applied to any Air Force instructional text. Analysis of a sample of Air Force instructional texts showed that all the sampled texts could be improved by revisions of this type.

C. We wrote software that can be used as a job aid by Air Force writers to revise any Air Force instructional text to improve its learnability; we

enclose two running programs.

II. BACKGROUND

The background for this project was the recent discovery that 72%-94% of the variance in the free recall of instructional text was accounted for by the frequency of cognitive gaps in the text (Britton, Van Dusen, Glynn, & Hemphill, 1990; Britton, Van Dusen, and Gulgoz, 1991). Frequency of cognitive gaps was measured by Kintsch's computational model of reading (Kintsch & van Dijk, 1978; Miller & Kintsch, 1980). This finding suggested three ways to improve learning from text in the Air Force:

A. Classify and select recruits for their abilities to successfully deal with the negative effects of cognitive gaps; this required developing and testing an individual differences model of the abilities needed for successfully dealing with cognitive gaps: this model is the topic of Section III, The Individual Differences Model.

B. Improve the learnability of Air Force texts by removing the gaps; this is the topic of Section IV, Revising Air Force Texts to Increase Learnability.

C. Develop computer job aids that Air Force writers can use to repair the gaps in Air Force instructional texts; this is the topic of Section V, The Computer Writing Aids.

What Are Cognitive Gaps in Text?

A cognitive gap can occur when a sentence's surface structure shares no ideas with the surfaces of any previous sentences. For example, there is a cognitive gap in the following text:

1. Title: Air War in the North, 1965.
2. First sentence: By the fall of 1964, Americans in both Saigon and Washington had begun to focus on Hanoi as the source of the continuing problem in the South.

The surface of the first sentence does not mention anything from the title: neither "Air" nor "War" nor "North" nor "1965". The result is that there is potentially a gap in the recruit's mind between the first sentence and the title. Unless the recruit does some extra mental work to fill the gap, the gap will remain present in the recruit's mental representation of the text. As in all cases where extra mental work needs to be done, some individuals will do the work and so fill the cognitive gap, but other individuals will not. (This text example is from the beginning of the Air Force text used in the present project: the whole text is included as Appendix A.)

Does Learning Decrease When Cognitive Gaps are Not Filled?

Two empirical studies show that less is recalled from naturally occurring texts that have more cognitive gaps in them (Britton, Van Dusen, Glynn, & Hemphill, 1990; Graves, Prenn, Earle, Thompson, & Johnson, 1991). In those studies, the correlations between the number of cognitive gaps in eight military history texts and learning from the eight texts were from -.85 to -.97, accounting for 72% to 94% of the variance--the more gaps there were, the less was recalled. Results from the experiments are shown in Table 1 (cognitive gaps were counted algorithmically by the Miller & Kintsch, 1980, computer program).

Table 1. Correlations Between Cognitive Gaps and Recall

Cognitive gaps counted by Miller & Kintsch (1980) program

	Learning	
Study	Immediate Recall	Delayed 24 Hours
Britton et al. (1990)	-.89*	-.85*
Graves et al. (1991)	-.97*	-.91*

*$p < .05$

How Does Filling Cognitive Gaps Increase Learning?

Filling cognitive gaps interconnects the different parts of the recruit's mental representation; in contrast, unfilled gaps lead to a mental representation that lacks connections between its parts. Interconnected mental representations are more memorable than unconnected representations, for several reasons.

First, comprehension of the text is increased when the learner is able to connect each sentence to the rest of his mental representation. Sentences that are unconnected to the rest of the learner's representation are not comprehended in the context of the rest of the representation.

Second, when the recruit comes to use his knowledge in a work situation, there is typically some cue in the situation that may or may not lead to retrieval of the needed information. For the retrieval to occur successfully there must be a mental pathway from the cue to the needed information. The more connected a piece of information is to other parts of the recruit's representation, the more likely there is to be a retrieval pathway from the cue to the needed piece of information. Pieces of information that are unconnected to the rest of the representation will not have retrieval pathways leading to them. But when the representation is interconnected, there is a heightened likelihood that a retrieval path will exist from the cue to the needed piece of information.

How are Cognitive Gaps Filled?

When a cognitive gap occurs between a sentence and the rest of the text, there are three possible outcomes: (a) the gap will be filled correctly because the learner does the correct mental activities, (b) the gap will be filled incorrectly because the learner does some mental activities, but they lead to an incorrect connection, or (c) the gap remains unfilled because the learner does not do extra mental activities that fill the gap.

What are the correct mental activities, the ones that maximize the chances that the gap will be filled correctly? The correct mental activity is to successfully infer how the sentence fits in to the text. This process of inferencing while reading instructional text is the core concept of this project.

As an illustration of the result of such an inferencing process, consider the earlier example from the Air Force Vietnam text. The learner who uses his inferencing ability could bridge the cognitive gap between the title and the first sentence by adding propositions to his mental representation of the text as follows (added propositions are shown in parentheses):

3. Title: Air War in the North (Vietnam), 1965
4. First sentence: By the end of 1964 (causing events in 1965), Americans in both Saigon and Washington had begun to focus on Hanoi (capital of North Vietnam) as the source of the continuing problem in the South.

If the reader infers the parenthesized propositions and adds them to his mental representation, then he is able to connect the first sentence to the title by the overlap between their ideas, namely the year and the country involved. And if the reader is able to do this sort of inferencing whenever it is needed throughout a text, then he will end up with an inter-connected representation of the text. This will increase learning for the reasons explained in the previous section.

How Does Successful Inferencing Work?

Occasionally, the recruit may be able to fill a cognitive gap by making a logical deduction that fits all the requirements of formal logic, but this is quite rare in most text. Usually the recruit must use non-deductive inference processes, such as inductive inference processes. Our central hypothesis was that recruits who have high levels of inference-making ability will be more able

to fill cognitive gaps in the text than recruits with low levels of the ability. Accordingly, we constructed three tests of inference-making ability and administered them to each recruit.

But more than inference-making ability alone is required to fill cognitive gaps. In the example, the recruit who is making inferences must also use as evidence certain domain-specific knowledge, such as the facts that the U.S. was engaged in a war with North Vietnam in 1964 and 1965, and that Hanoi is the capital of North Vietnam. But he can only use this knowledge if he has it. Accordingly, a second part of our hypothesis was that having high levels of such domain-specific prior knowledge will help the recruits to fill cognitive gaps. We used three tests of domain-specific prior knowledge.

We have mentioned several prior knowledge elements that eventually got used in this example. But of course the learner may also have considered other prior knowledge elements that eventually turned out not to be useful for bridging this particular gap. How does he settle on which pieces of knowledge to use, and how does he decide what propositions to add, and in which locations in his mental representation?

Our hypothesis is that when he tries to fit various possible elements of prior knowledge into the text, he comes up with various possible propositions to add, which he then tries out until he finds one or more that successfully fills the gap. To try out any such possibility, (a) he must hold in mind several elements, including (in the example) at least the title, the first sentence, and the prior knowledge elements he is currently trying, and (b) he must then perform the mental operations needed to draw conclusions from the elements, try out candidate propositions, and judge whether they fill the gap. If they successfully fill the gap, then he is finished, but if not, he must try other possibilities until he finds a solution that satisfies him, or else gives up.

According to this description of the inferencing process, a successful outcome often requires holding several things in mind while performing various mental operations on them. The component of the human cognitive system that is specialized for holding several things in mind while mentally manipulating them is working memory (Baddeley & Hitch, 1974; Kyllonen, 1991) and so a third part of our hypothesis was that having high levels of working memory ability will help recruits to fill cognitive gaps. We used two tests of this ability. In summary, according to our description of the inference-making process so far, inference-making ability is the main player, but it must be supported by domain-specific prior knowledge and working memory ability.

A remaining question is: what triggers the inferencing machinery into action? If the learner's cognitive system senses that a cognitive gap has occurred, this could be the trigger. But we know that many readers fail to notice cognitive gaps in text, even such obvious gaps as direct contradictions in simple stories (Ackerman, 1986; Schmidt & Paris, 1983). And those studies also show that the ability to spot that a cognitive gap has occurred differs greatly among individuals. So the final part of our hypothesis was that having high levels of gap spotting ability will help the recruit to trigger the inferencing machinery for filling cognitive gaps; we used two tests of this ability.

The model of the inferencing process that we developed is based on those four abilities: at its center is inference-making ability; supporting roles are played by prior knowledge and working memory ability; and the inferencing machinery is triggered by gap-spotting ability.

It may be helpful to clarify our use of the terms inferencing process, inferencing machinery, and inference-making ability. "Inferencing process" is the broadest term; it refers to the whole of the mental process that successfully

fills a cognitive gap. By "inferencing machinery", we mean a subset of the inferencing process, namely the moving parts of the system: inference-making ability proper in interaction with domain-specific prior knowledge, performing operations in the arena provided by the capacity of working memory. Inference-making ability proper refers to the ability to do deductive and inductive logical operations with natural language materials; performing such logical operations necessarily makes use of both linguistic prior knowledge (such as the meaning of words and the parsing of syntax) as well as whatever general world knowledge is needed to interpret the words and syntax (such as the existence of males and females, the social structure of possession, the operation of doors, etc.) So the linguistic and general prior knowledge needed to do logical operations with natural language is tested within our inference-making tests. It must be distinguished from domain-specific prior knowledge about the topic of the text.

III. THE INDIVIDUAL DIFFERENCES MODEL

We have described a cognitive process model that specifies the causal relationships among the four abilities and learning. To test it, the process model was translated into a structural model and used to predict learning from a 1000-word Air Force text. This structural model, called Model 1, is shown in Figure 1.

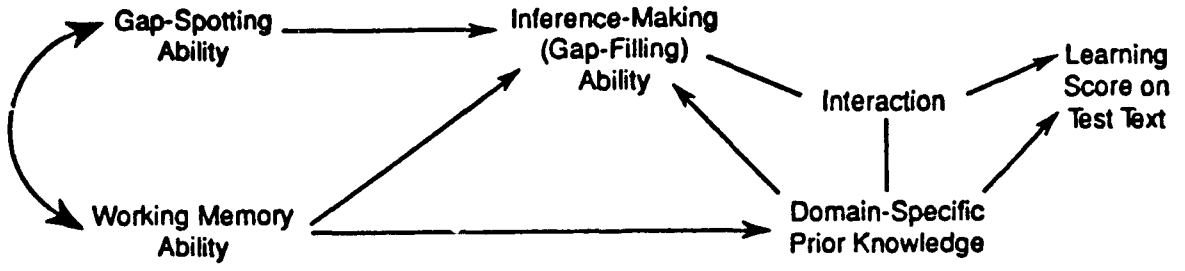

Figure 1. Model 1: Relating Learning to Four Individual Differences Abilities

In addition to measuring learning, we also tested the inferences that

recruits made from the Air Force text. We did this because, besides determining learning, our hypothesized cognitive process should also determine whether recruits made correct inferences from the Air Force text from which they were learning. Also, recruits who made more correct inferences from the Air Force text should learn more from it. These ideas led us to propose Model 2: it includes our measure of the inferences that recruits made from the Air Force text, and puts those inferences on the causal path to learning.

According to our hypotheses for Model 1, the learning score (shown at the right of Figure 1) is determined by the interconnectedness of the recruit's text representation; in texts with gaps, some of this connectedness must be manufactured by the recruit, as in Examples 3 and 4 above, by adding propositions that are constructed by making inferences using prior knowledge. Making inferences using prior knowledge was represented in the model as the interaction of the individual's inference-making ability with his domain-specific prior knowledge. In Figure 1, it is labeled "interaction"; to show its origin, it is connected by lines to "inference-making ability" and "domain-specific prior knowledge". This interaction has a direct effect on learning, as shown by the arrow leading to "learning".

In addition, recruits with domain-specific prior knowledge about the topic of the text may be able to use that knowledge directly to answer questions on the learning test. This is represented in Figure 1 as the direct effect of domain-specific prior knowledge on the learning score.

Working backwards along the causal sequence toward the left, working memory ability is indirectly related to learning, with its effects completely mediated by its relations to inference-making and prior knowledge, and through them, its effects on their interaction. Finally, gap-spotting ability is indirectly related to learning, with its effects completely mediated by its effects on

inference-making ability; one is more likely to successfully deploy one's inference-making ability if one's cognitive system has sensed that an inference is called for, because a gap has been spotted.

Model 2 is shown in Figure 2. It adds the measure of the inferences made from the test text. Our process model of the inference-making process proposes that the inferences made on the test text should be determined by inference-making ability, prior knowledge, and working memory ability, and that learning should be determined in part by inferences made on the test text and in part by the interaction between inference-making ability and prior knowledge, as well as by domain-specific prior knowledge.

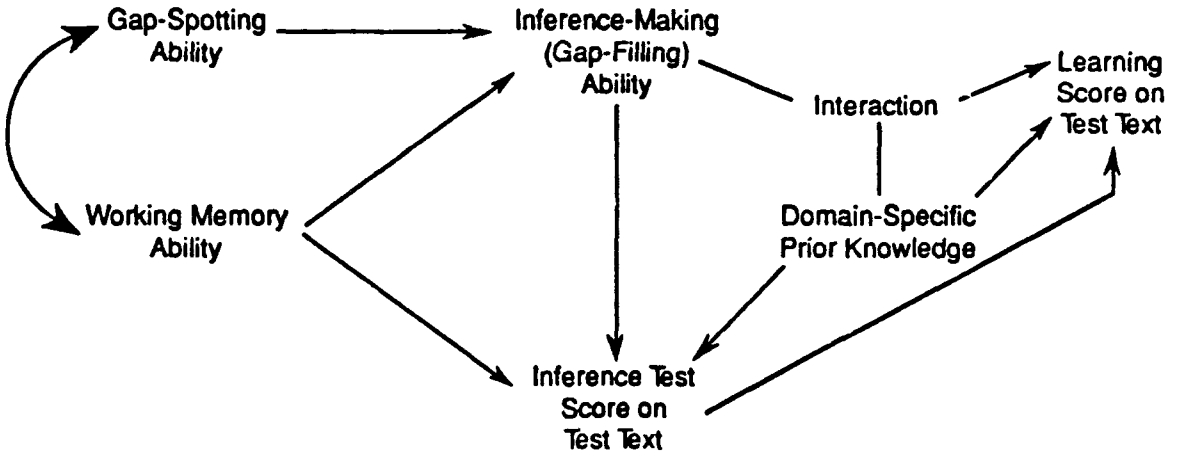

Figure 2. Model 2: Relating Inferences from Text and Learning to Four Individual Differences Abilities

Method

Individual Difference Tests

Gap Spotting Ability. Stories with contradictions were presented one sentence at a time, followed by a question asking whether the story had anything in it which did not make sense. A sample item is shown in Figure 3. Two tests were used, including 13 stories on adult topics from Ackermann, 1986, and 9

stories on children's topics from Schmidt and Paris, 1988 .

Nancy was working around the house.

Nancy washed five of her shirts and wanted to dry them on low heat.

She put the shirts in the dryer.

The cat was on the porch.

The cat had a bird in its mouth.

The shirts were hung out on the line.

Test item:

Did you notice anything about the story that did not make sense?

D (NO) L (YES)

Figure 3. Sample Gap Spotting Item

Inference-Making Ability. Three tests were used. Pronoun inference items were 19 brief stories with a pronoun for which the referent had to be inferred (from Oakhill, 1986). A sample item is shown in Figure 4.

Joseph and Fred met at the school music club.

Joseph sold his guitar to Fred

because he wanted to play.

The members seemed to enjoy themselves and

he wanted to learn to play

so that he could join in the fun.

_____ wanted to learn to play.

D--Fred L--Joseph

Figure 4. Sample Pronoun Inference Item

Single-sentence inference items were 32 items for which a single sentence gave enough information to make the inference necessary to answer the question (from Vosniadou-Papanicolas, 1982). A sample item is shown in Figure 5.

Today Dave has failed to cage the parrot.

Test item:

Parrot

D--caged L--free

Figure 5. Sample Single-Sentence Inference Item

Story inference items were six long stories from ten to thirty sentences, each of which had several questions requiring inferences to answer correctly (from Singer, 1988). Figure 6 shows a sample item.

David and Jan went on a picnic.

David split some logs to make a fire.

They sat down and ate their sandwiches.

Jan accidentally dropped the wine she was pouring.

A wet stain spread over the blanket.

Suddenly hundreds of mosquitoes attacked.

David quickly put out the fire

They raced from the woods.

Jan turned on the ignition key.

The motor hummed as they drove off.

They decided to go to a restaurant next time.

Test items:

Did the wine spill?

D L

(NO) (YES)

Did the car start?

D L

(NO) (YES)

Did the ants attack?

D L

(NO) (YES)

Figure 6. Sample Story Inference Item

Domain-Specific Prior Knowledge. Multiple choice items on the domain of the text were used. One test included 3 items on the topic of the text but not

mentioned in the text (from an Air Force document not read by the recruits), another test included 3 items on topics involved in the Vietnam War but not mentioned in the text (Figure 7 shows a sample item from this test), and the third test included 8 items on historical events that occurred at the same time as the Vietnam War, but were not directly concerned with the Vietnam War, nor mentioned in the text.

Da Nang was:

1. President of South Vietnam at the time of South Vietnam's surrender to North Vietnam.
2. a river separating Vietnam from Thailand.
3. the leader of the South Vietnamese rebels.
4. the site of a large North Vietnamese military hospital.
5. a city in South Vietnam that contained several military bases.

Figure 7. Sample Domain-Specific Prior Knowledge Item

Working Memory Ability. Figure 8 shows a sample item (from Kyllonen, 1991). A pair of terms were presented in a sentence, and then that sentence was removed so that the recruits had to hold the terms in mind. Then another such sentence was presented and removed. Then a final sentence instructed the recruits to mentally manipulate the information from the earlier sentences in a specific way. Finally, 8 choices were presented, only one of which was the correct result of the prescribed mental manipulation. Thirty-seven such items were presented.

The bird comes before the dog.

The rug comes after the chair.

The ANIMALS do not come after the FURNITURE.

Test items:

Which of these is the correct answer?

1.	bird	dog	rug	chair
2.	bird	dog	chair	rug
3.	dog	bird	rug	chair
4.	dog	bird	chair	rug
5.	rug	chair	bird	dog
6.	rug	chair	dog	bird
7.	chair	rug	bird	dog
8.	chair	rug	dog	bird

Figure 8. Sample Working Memory Test Item

To-Be-Learned Text and Tests on It

Air Force Instructional Text. The text was from U.S. Air Power: Key to Deterrence (U.S. Air Force Reserve Officers' Training Corps, 1985). the passage was a self-contained, complete section, taken from Chapter 34 of Ehrhart (1978). Its length was 1030 words and its Coleman-Liau (1975) readability formula score was grade level 12.5. The text is shown in Appendix A. (All texts and tests used in this experiment are available from the author on request.)

Learning Test. Our goal for the learning test was to assess the degree of correspondence between the shape of the learners' mental representations and the shape of the mental representation intended by (a) the text's author, and (b)

qualified subject matter experts.

The intended mental representation was provided by the author of the original text (a professor in the Department of History at the U.S. Air Force Academy). An independent delineation of the intended mental representation was provided by seven subject-matter experts, including a former U.S. ambassador, two academic military historians and one amateur, a current U.S. Air Force major, and two high-level former military employees.

To measure the shape of the mental representations, we chose 12 important terms about the text and constructed all possible pairs of them, 66 in all. Then we asked each subject to rate each pair for relatedness on a 7-point scale. The result was a column of 66 ratings for each subject. Degree of relatedness is a measure of proximity, so maps can be made of the set of ratings to depict psychological distance relations.

We interpret the recruits' ratings as a quantitative representation of the recruits' ideas about the relationships among the 12 terms. The ratings of the author of the original text represent the relationships he intended his text to communicate about these 12 terms. The 7 subject-matter experts, as did the author, read only the original text; the experts, like the author, had extensive prior knowledge about the subject-matter of the text, so their ratings can be interpreted as a combination of (a) the information in the text and (b) their relevant prior knowledge. Probably the only information the experts lack that Colonel Ehrhart has is the author's introspective access to his intentions in writing the text.

The representations of each of the U.S. Air Force recruits can then be compared to the intended representations: those of the author and the 7 subject-matter experts. If the correspondence is high, we can conclude that the recruits got the intended mental representation. The degree of correspondence was

evaluated by calculating a correlation coefficient between the column of 66 numbers provided by each recruit and those provided by the author or experts.

Author. The author was Colonel (then Major) Robert C. Ehrhart, then of the Department of History of the U.S. Air Force Academy (now retired). He read only the original version. He was given instructions by phone by Bruce K. Britton and took the test by mail.

Experts. Seven local experts were contacted, and all agreed to participate. Ambassador Hillenbrandt was U.S. Ambassador to West Germany and Hungary and Assistant Secretary of State for European Affairs. The military historians were professors in the University of Georgia History Department. Major Eunice was commandant of the local U.S. Air Force Reserve Officers' Training Corps (ROTC) and instructor for the ROTC course for which these materials were used. Other experts included two former military employees and an amateur military historian. All experts read only the original version and completed the test in their office or at home. The first 4 were tested by Bruce K. Britton and the remaining 3 were tested by Anne Reynolds.

The 12 terms were selected by Bruce K. Britton with Anne Reynolds to represent important topics of the text. To clarify the character of the test, the content of the passage is described briefly in Figure 9, with the terms used in the test underlined. (Figure 9 was not shown to the subjects; it is only used in this report as an expository aid):

The texts began by describing some members of the Johnson Administration, including President Johnson, who had civilian advisers, including Robert McNamara and Maxwell Taylor, as well as military advisors. The military advisers proposed the military strategy, which was (roughly) to bomb North Vietnam very heavily. The civilian advisers proposed instead the graduated response strategy, which was to bomb North Vietnam a little and then pause to see if that had "broken their will"; if it hadn't, the bombing would be escalated gradually. Since the focus was on breaking the North Vietnamese will, this was described in the passage as a psychological strategy. Johnson chose the graduated response strategy, and the resulting operation was code-named Rolling Thunder. Success and failure could be attributed to various persons, policies, actions, and consequences in the passage.

Figure 9. Terms Used in Test of Learning

All possible pairs of these terms were presented in random order, each with a 7-point scale ranging from very closely related (1) to very distantly related (7). The recruits saw each pair of terms on the computer screen and responded by pushing the keys marked with the numbers 1 to 7. The author and the experts took a paper-and-pencil version of the same test.

Inference Test. This 32-item test can be described best by example. The first inference question dealt with the inference called for to connect the second sentence of the passage to the first one:

First sentence: By the fall of 1964, Americans in both Saigon and Washington had begun to focus on Hanoi as the source of the continuing problem in the South.

Second sentence: As frustration mounted over the inability of the ARVN to defeat the enemy in the field, pressure to strike directly at North Vietnam began to build.

In order to connnect these two sentences as the author intended, the reader must make the inference that "the inability of the ARVN to defeat the enemy in the field" *refers to the same thing as* "the continuing problem in the South". (See Section III of this report for the basis of this claim.) The multiple-choice question that tested whether this inference was available was this:

15. The inability of the South Vietnamese Army (ARVN) to defeat the enemy in the field was regarded as:
    a. the aerial extension of the war.
    b. a result of not being familiar with the terrain.
    c. lack of carefully planned strategies.
    d. the air war in the North.
    e. the continuing problem in the South (correct response).

Procedure

Order of Tests. First the Air Force test text was presented, followed by its learning test and the test on inferences from it, followed by the gap spotting and inference-making items intermixed, followed by the prior knowledge tests, followed by the working memory tests.

Learning Test Procedure. Each recruit read the text and took the test on a Zenith computer. The recruits pushed the space bar to present each sentence, with the previous text remaining on the screen until a page was completed; then all the text on that screen disappeared, and the new text started at the top of

the screen. Once the text was completed, it was followed by the test. The experts read the text on paper and then took the test.

Learning Test Instructions. The recruits were instructed to read the text carefully. For the test, they were told to press a button for each pair to indicate how related they thought the pair had been in the text they had just read.

The experts were asked to read the text and were told that they would be asked to fill out a questionnaire on it. After they finished reading, they were shown the test. They were told that we were interested in what the text was trying to express and that they should try to set aside their own personal views on the historical events described in the text.

Analyses

Analyses of the Learning Test. Correlational analyses were conducted of the ratings on the 66 pairs, first comparing the author's ratings to each expert and to the combined group of experts, to measure the agreement among them. Since the results of this first analysis showed a very high level of agreement among the author and experts, the remaining analyses used the combined experts' ratings as the criterion against which each recruit's learning was compared. Second, similar correlational analyses on the 66 pairs were used to compare each recruit to the combined group of experts. The result was a single correlation for each recruit, which measured his correspondence with the experts' structure. This correlation was then transformed with Fisher's z, and used as the measure of learning for each recruit.

Structural Models. The cognitive process model on which this investigation is based is described in the section on How Does Successful Inferencing Work. This process model suggests a structural model in which (a) the causal path to learning includes an interaction between inference-making ability and domain-

specific prior knowledge, and (b) there are no direct connections between gap spotting ability and learning, nor between working memory and learning. To address the interaction, we used residual centering (Lance, 1988) to test all possible interactions. Only the interaction between inference-making ability and domain-specific prior knowledge was significant. Therefore, we did not consider any models with other interactions. Condition 9 and Condition 10 tests were conducted for both path models (James, Muliak, & Brett, 1982). Condition 9 tests, often called "omitted parameters tests", began with a model including all possible paths, and compared it with our hypothesized model. If the paths not in our model had coefficients not significantly different from zero, the Condition 9 tests support our model. Results of the Condition 9 tests of the models were consistent with our hypotheses on the absence of direct causal connections between gap spotting ability and learning, and between inference-making ability and learning. Condition 10 tests were conducted to assess the fit between the models and the empirical data. Condition 10 tests examine the significance of the paths that were included in the models. Since the weights on those paths were significantly different from zero, the Condition 10 tests supported our models. The results are presented in Figures 10 and 11.

A principal components analysis was conducted to confirm that we had a satisfactory measurement model (Bollen, 1989). Given a satisfaction measurement model, heirarchical confirmatory factor analysis is an appropriate method for confirmatory analysis (Marsh & Hocevar, 1988). Heirarchical confirmatory factor analysis was implemented using LISREL VI software (Joreskog & Sorbom, 1984). The results are presented in Figure 12.

Results

Condition 9 and Condition 10 tests were successful for both models. Figure 10 shows Model 1. The coefficients on the paths (standardized regression

weights) were significant and in accord with our hypotheses. Gap-spotting ability had an indirect effect on learning that was completely mediated by inference-making ability. Working memory ability had an indirect effect on learning that was completely mediated by inference-making ability and prior knowledge. The interaction between inference-making ability and prior knowledge was directly related to learning, as was prior knowledge. Table 2 shows the form of the interaction between inference-making ability and prior knowledge. The interaction was such that recruits who had high scores on both learned more than recruits who had high scores on only one or neither. Model 1 accounts for 33% of the variance in learning.

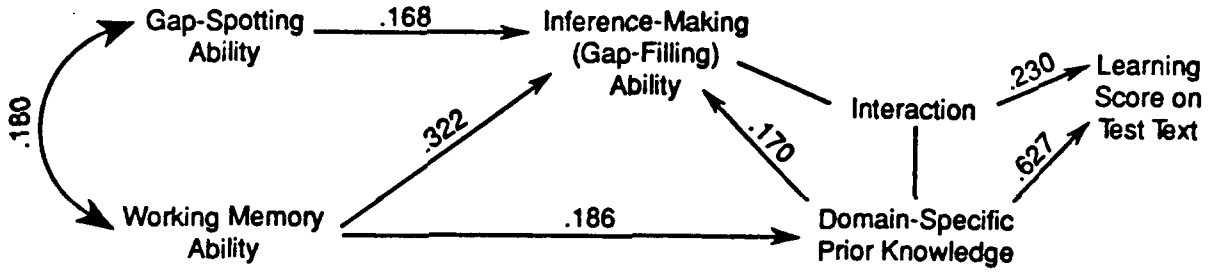

Figure 10. Structural Model 1

Table 2. Effects of Interaction Between Inference-Making Ability and Prior Knowledge on Learning Scores

Tabled numbers are mean learning scores (SE in parentheses)

		Inference-Making Ability	
		Low[a]	High
Domain-Specific Prior Knowledge	Low	.009 (.024)	.035 (.031)
	High	.195 (.049)	.438 (.061)

[a]Low and High refer to median splits

Figure 11 shows Model 2, to which we had added the measure of the inferences that were made from the test text itself. As expected, the inferences made from the test text itself were a major contributor to learning, in accordance with our hypotheses: Inference-making ability, prior knowledge and working memory all contributed directly to actually making correct inferences in the test text, while gap-spotting ability had an indirect effect, completely mediated by inference-making ability. Model 2 accounts for 54% of the variance in learning.

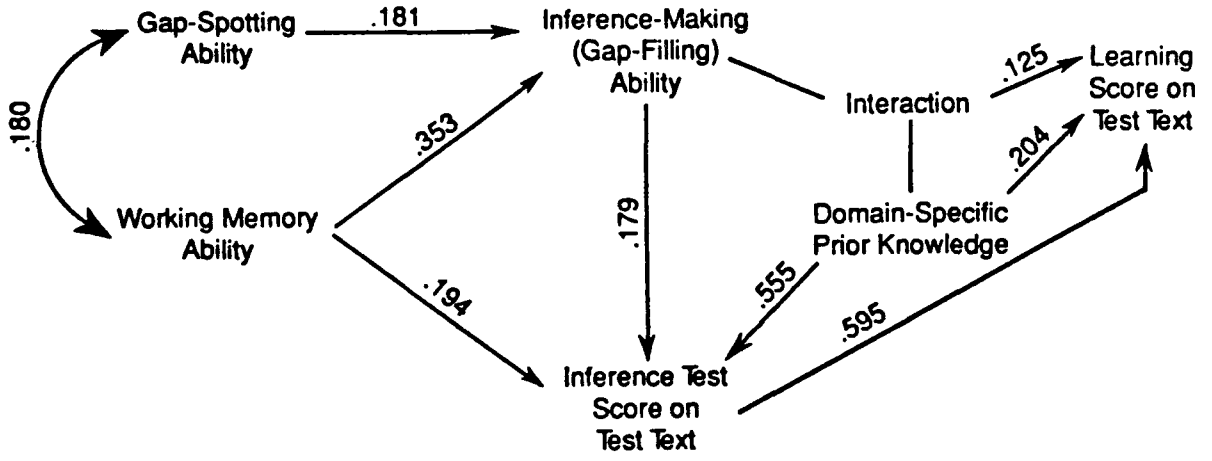

Figure 11. Structural Model 2

Figure 12 shows the LISREL model of Model 1. Use of LISREL is justified by our confirmatory stance, which is in turn justified by the match between our cognitive process model and the structural models of Figures 10 and 11. The LISREL model had a Joreskog & Sorbom (1984) goodness of fit index of .97. Alternative goodness of fit measures included Root Mean Square error of .034, Bentler's (1990) normed fit index of .951, and Tucker-Lewis index of .904.

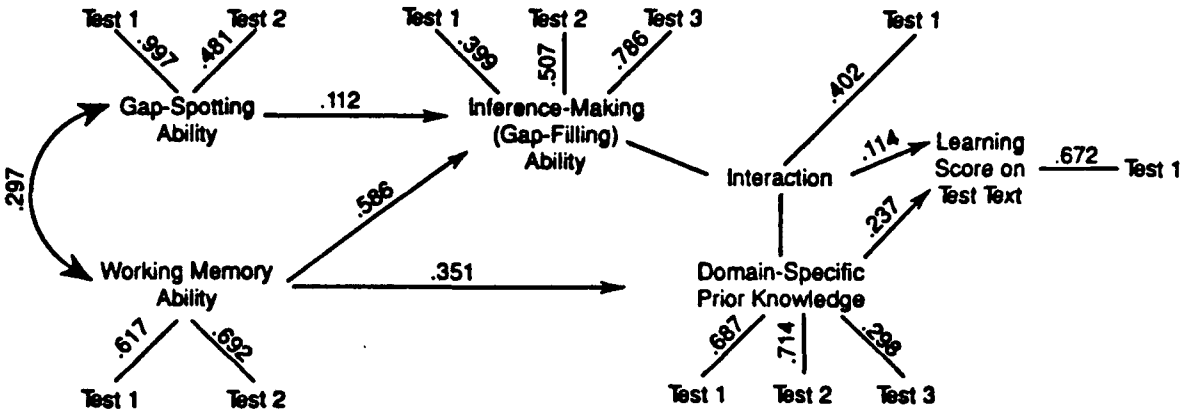

Figure 12. LISREL Model of Structural Model 1

Discussion

The success of the structural models implies that the independent variables are causally connected to learning as predicted by the cognitive process model. The relationships between gap spotting, working memory, inference-making ability, prior knowledge, and learning have important implications for the effects of individual differences on recruits' learning of Air Force texts. The next step was to use our understanding of these abilities to improve the learnability of an Air Force text, by reducing the demands that the text made on the individual's abilities.

IV. REVISING AN AIR FORCE TEXT TO INCREASE ITS LEARNABILITY

Our goal in this part of the project was to test a computational method for improving instructional text. The computational method was based on two things: (a) the individual differences variables identified in Section III, and (b) Kintsch's theory of reading (Kintsch & van Dijk, 1978) as implemented in his

computer program (Miller & Kintsch, 1980). Kintsch's program simulates the way in which the human reader constructs a mental representation of the text. In the simulation, first the program inputs the first sentence's propositions; then it inputs the second sentence's propositions and looks for coherence of the propositions of the second sentence with those of the first sentence. Coherence is found when the second sentence contains an idea previously mentioned in the first sentence. Then the program inputs the third sentence, looks for coherence, and so forth; the process continues until the end of the text.

What happens when coherence fails to be established because a new sentence does not mention an idea from a previous sentence? What happens is that there is a cognitive gap in the text. As pointed out in the Section on Background, the human reader may try to make an inference to link the new sentence with some idea from a previous sentence. If the reader succeeds in making a coherence-establishing inference, then coherence is maintained. However, if he or she fails, then coherence is not maintained, with the result that the reader's mental representation of the text is not coherent. The reader's mental activities while making such inferences are determined by four abilities, according to our model: gap-spotting, inference-making, working memory, and prior knowledge.

In contrast to a human reader, when Kintsch's computer program encounters the call for an inference, it simply indicates that it is unable to continue. This is because the program is not equipped to make inferences.

Our experiment was concerned with cases in which such calls for inferences occurred. We believed that texts for which coherence has been successfully established are more learnable, so we expected that texts with fewer inference calls would be more learnable. To identify the inference calls, we ran the texts through Kintsch's program.

The background for our study was a previous study in which we used the

Kintsch program to predict learning from texts (Britton, Van Dusen, Glynn, & Hemphill, 1990); the results are reported in Table 1. However, this was an entirely correlational study.

The obvious next step was to conduct the corresponding experimental study. Many investigators had conducted experimental studies in which they began with unclear or ambiguous texts and then tried to improve them; 62 such studies were reviewed by Britton, Gulgoz, and Glynn (in press). Our experimental study was innovative in that we used a cognitive process model to specify theory-based modifications rather than intuitive or global ones. In this way, we tried to bridge the gap between computational psychological theories and instructional practice. A full report of the present study is in Britton and Gulgoz (1991).

In our study, we used our individual differences variables, as embodied in the principles of the Kintsch program, to analyze an Air Force instructional text as we had in the Britton et al. (1990) study, to identify the text locations at which the reader would have to make an inference to establish coherence. We found 40 such locations. Then we repaired the text at each of those locations by inserting the required inference into the text itself.

In this way, we produced a principled revision of the text in which the "principles" were the principles underlying the Kintsch program. We expected that recruits who were reading the principled revision would have a more coherent memory representation than would recruits reading the original text because we increased the coherence of the text in the principled revision by inserting the missing inferences. We expected that this would lead to better memory for the information in the text because recruits would have more links between different parts of their memory representation and could use those links to get to more parts of the text's representation.

Method

Principled Revision

The principles used in the revision are illustrated by example, starting at the beginning of the text:

1. Air War in the North, 1965
2. By the Fall of 1964, Americans in both Saigon and Washington had begun to focus on Hanoi as the source of the continuing problem in the South.

Sentence 2 has no argument overlap with Sentence 1, so the Kintsch program cannot process it. An alert, active, and knowledgeable recruit, however, can infer a link between the late part of "1964" causing something in the year "1965", which is mentioned in the title, or a link between "Hanoi", capital of North Vietnam, and the "North" of North Vietnam referred to in the title. These inferrable links are shown in parentheses in Sentences 3 and 4:

3. Air War in the North (Vietnam), 1965
4. By the Fall of 1964 (causing events in 1965), American in both Saigon and Washington had begun to focus on Hanoi (capital of North Vietnam) as the source of the continuing problem in the South.

The words in parentheses represent propositions that the recruit could insert into his or her mental representation of the text. When these propositions are inserted into the proposition list that is the input to the Kintsch program, the program can easily establish coherence between the sentences by idea overlap. In the principled revision, we incorporated such changes into the text itself:

5. Air War in North Vietnam, 1965
6. By the beginning of 1965, Americans in both Saigon and Washington had begun to focus on Hanoi, capital of North Vietnam, as the source of the continuing problem in the South.

The principle used in these sentences can be briefly stated as Principle 1: Make the recruit's job easier by rewriting the sentence so that it repeats, from the previous sentence, the linking word to which it should be linked. We have already illustrated this in Sentences 5 and 6. The derivation of this principle from Kintsch's model is straightforward. His model finds coherence when an idea is repeated from one sentence to the next; the most obvious way to repeat an idea is to repeat the word that refers to the idea. In Kintsch's theory, this is referred to as "argument overlap"; each proposition has one or more arguments (ideas) in it, and overlap occurs when the same argument is present in a later proposition. In terms of our individual differences model, inserting the link into the text obviates the necessity to make the inference.

We applied Principle 1 extensively throughout the original text. For example, in the second full sentence,, the phrase, "the inability of the ARVN to defeat the enemy in the field" is actually the same thing as the first sentence's "the continuing problem in the South". That is, "the inability of the ARVN to defeat the enemy in the field" is "the continuing problem in the South". However, we expect, as does the Kintsch program, that many recruits miss this connection because no content words are repeated across the two sentences. In the principled revision, we changed the first sentence to read "the continuing war in South Vietnam" and the second sentence to read "the South Vietnamese Army was losing the ground war against North Vietnam" (italics added in both phrases for purposes of illustration). We expected that some of the recruits who would otherwise have missed it would make this link and incorporate it into their representation, creating a more correct and complete representation, i.e., increasing their comprehension of the text. Because these links can be followed during free recall, these recruits will also have established additional retrieval routes for themselves.

The alert reader will have noticed that we made some additional changes in those phrases. These changes were motivated by a corollary of Principle 1: Whenever the same concept appears in the text, the same term should be used for it. We applied this corollary because we found that the original version involved the use of many different terms for the same concepts: 12 different terms for bombing attacks (i.e., air war, strike directly, aerial extension, bombing attacks, campaign, air campaign, attacks, bombing, American power, bombing operations, air raids, air power); 23 different terms for American officials; 15 different terms for North Vietnam; and 6 different terms for South Vietnam.

We applied Principle 1 to change these different terms to the same terms whenever possible. In the first and second sentences, this motivated the change from "Americans" in the original to "American officials" in the principled revision; "Saigon" to "South Vietnam"; "South" to "South Vietnam"; and "ARVN" to "South Vietnamese Army".

Principle 2 is to make the recruit's job easier by arranging the parts of each sentence so that (a) the recruit first encounters the old part of the sentence, which specifies where that sentence is to be connected to the rest of his or her mental representation; and (b) the recruit next encounters the new part of the sentence, which indicates what new information to add to the previously specified location in his or her mental representation.

The advantage of this arrangement can be seen by considering the alternative: Suppose the new part of the sentence appears first, without the recruit knowing where in his or her representation to put it, followed sometime later on by the old part, which indicates where in the recruit's mental representation the new part is to be placed. This arrangement imposes an additional load on working memory; the new part must be held in working memory

while the old part is being processed to determine where the new part should be put. This extra load is obviated by following Principle 2. So in our individual differences model, the given-new revision principle is related to the working memory component.

In the Kintsch model, the warrant of Principle 2 is the leading-edge strategy, which implements this old-new contract (Haviland & Clark, 1974). We can use the old-new principle because nearly every sentence has an old part, sometimes called the "given part" (by which it is connected to the previous text), and a new part, which adds the new information by which the text's representation grows. (An exception is the title of a text, which, as in this example, is all new information for that text.)

The first example of a text revision based on the given-new principle in this text is in the second sentence with regard to the first sentence. First we must notice that the second sentence is actually a compound sentence, composed of two separate complete sentences joined by the conjunction "As", so the units involved in this example can be written as follows:

7. By the Fall of 1964, Americans in both Saigon and Washington had begun to focus on Hanoi as the source of the continuing problem in the South.
8. As
9. frustration mounted over the inability of the ARVN to defeat the enemy in the field,
10. pressure to strike directly at North Vietnam began to build.

Should we rearrange Sentence 9 in such a way that (a) the first thing mentioned is the given information, allowing the learner to move his or mental pointer in the mental representation at which the rest of the sentence should be added, and (b) the second thing mentioned is the to-be-added information?

Our first step must be to identify the old and the new information. In

Sentence 9, it is clear that "frustration mounted" has never been mentioned before, so it must be new information. In addition, it is evident in the example under Principle 1 that "the inability of the ARVN to defeat the enemy in the field" is actually a repetition of an old idea from Sentence 7 (i.e., "the continuing problem in the South"). Therefore, the new information must be "frustration mounted" and the old information must be "the inability of the ARVN to defeat the enemy in the field" (whereas "over" relates the old to the new by the relation of causality).

In Sentence 9 we can see that the new information comes first and that the old information comes second. (This happens in 20 cases in the original version.) Thus, the recruit encounters the to-be-added information before he or she knows where to add it. Principle 2 states that the recruit's job would be easier (a) if the recruit first knew where to connect the information to his or her existing mental representation and (b) once the recruit's "mental pointer" were moved to the correct location, if he or she received the next piece of information needed in the learning process: the information to be added to that location in his or her mental representation. To help the recruit, we first reversed the order of Sentence 9 to put the old information first:

11. The inability of the ARVN to defeat the enemy in the field (old) caused frustration to mount (new).

From this point, the route to the final principled revision is straightforward. Because "ARVN" is one of the many different terms referring to the South Vietnamese, but because this is not specified anywhere in the original version, we changed the term to the "South Vietnamese Army", on the basis of the corollary of Principle 1. Under Principle 1, we have shown why "the inability of the ARVN to defeat the enemy in the field" was changed to "the South Vietnamese Army was losing the ground war against North Vietnam", so we proceed

from Sentence 11 to the following:

12. The South Vietnamese Army was losing the ground war against North Vietnam (old) and this caused frustration (new).

The reasons for adding the phrase "among the American officials" is explained later under Principle 3.

As mentioned earlier, the warrant for Principle 2 in Kintsch's model is his leading-edge strategy. The leading-edge strategy is the algorithm that is used in the Miller and Kintsch (1980) model to choose which earlier propositions to hold over in working memory from one input cycle to the next. These held-over propositions are the ones that the program examines first when looking for coherence links between the new sentence and earlier propositions. The leading-edge algorithm operates by first examining the ideas in each newly input sentence, looking for the first idea in the sentence that repeats an idea that has previously been mentioned. This corresponds to the previously given idea; the algorithm then puts that idea at the top of the tree structure that it is creating for this sentence. Then, as it approaches the end of its operation for this sentence, the leading-edge algorithm hooks up the latest ideas in the sentence to the tree structure. These latest ideas are likely to include the new ideas if the sentence takes the form of the canonical order of given followed by new. Finally, the leading-edge algorithm chooses for holdover those ideas at the top of its tree structure, which are likely to include the given ideas, and the latest ideas in the tree structure, which are likely to include the new ideas. Thus, the leading-edge algorithm has a built-in bias to choose propositions likely to include the given ideas and propositions likely to include the new. In developing their model, Kintsch and his associates tried a number of holdover strategies; the leading-edge strategy seemed to work the best in simulating readers' performance. This may be because it mimics the given-new principle, as

do authors and readers: authors when they write and readers when they read.

Principle 3 is to make the recruit's job easier by making explicit any important implicit references; that is, when a concept that is needed later is referred to implicitly, refer to it explicitly if the recruit may otherwise miss it.

We included this principle because when we run texts through the Kintsch program we often find a conceptual gap as a result of a later sentence referring explicitly to something that is only implicit in an earlier sentence. Because sometimes human readers do not make explicit what is implicit, they too may miss such links. By making such implicit ideas explicit in the text, we implement another form of inference reduction. For example, Sentence 9 refers to "frustration" but does not state who is frustrated; because only animate beings can be frustrated, "frustration" must refer to some person or persons. Passive, unmotivated, confused, or overburdened recruits may fail to figure out that it must be the "Americans" of Sentence 7 who are frustrated. The Americans are important players in this text, so we insert the needed information into the text. ("Americans" was changed to "American officials" by the corollary of Principle 1.) Similarly, the link between "frustration" and "pressure" is implicit in the original version but is made explicit in the third sentence of the principled revision.

We have provided several examples of the text's problems and our repairs. It is also important to delineate the conceptual distinction between (a) how we identified the locations of the inference calls, and (b) how we repaired them. As mentioned previously, we identified the locations of inference calls by using the principles of the Kintsch program. It is important to realize that the repair also involved principles, but these principles were not the ones directly implemented in the Kintsch program, because the Kintsch program is a model of

reading, not a model of how to repair text.

Instead, the principles we used to repair the text were derived from the Kintsch theory in another way: Our repair principles were chosen to coordinate with Kintsch's theory, in that they modified the text so that the Kintsch program could process it. That is, whenever the coherence could not be established without an inference, the Kintsch program could not continue its processing; it stopped. When our revision supplied the inference, the Kintsch program could then proceed to process the text until the next inference call was encountred. Thus, our repair principles were chosen to supply what the Kintsch program needed to continue processing; it is in this sense that our repair principles were derived from the Kintsch program. Our repair principles were also derived from our individual differences model, in that our repairs supplied to-be-inferred propositions which would otherwise have been absent because recruits were missing one or more of the four abilities that enabled them to make such inferences.

This method of repairing the text raises the question of what we did when the text allowed for more than one coherence-establishing inference: Which of the possible revisions did we make? Our practice was to follow the author's intention, which we determined from the text. For example, we believed that in Sentences 1 and 2 the author intended the reader to link "the Fall of 1964" with the year "1965) in the title and to link "Hanoi", through its identity as the capital of North Vietnam, with "North" in the title, through its identify as North Vietnam. That is, we believed that the author had these links in his mind and intended them to be in the reader's mind. So whenever we had a choice about which possible inference to put in the text, we resolved it by trying to determine what the author intended and implementing that. For more details on the principles and their application, see Britton & Gulgoz (1991), and Gulgoz (1989). The principled revision is presented in Appendix B.

No-Protocol-Guided Revision

The principled revision was not modified after collecting data from subjects. That is, we did not write the revision, collect data on its learnability, and then change the revision. Instead, we wrote the revision first and then tested it only in the experiments reported in this study.

Procedure

This study was conducted at the same time as the individual differences study reported in Section III. A total of 213 recruits were randomly assigned to read the Principled Revision.

Results

The groups reading the Original Text and the Principled Revision did not differ reliably on any of the individual differences variables, nor on AFQT.

The average correlation between the average experts' structure and the individual structures of subjects who read the Original Text was .15, and the corresponding correlation for the Principled Revision was .30. These differ reliably $t(211) = 4.64$, $p < .0001$. (This test was conducted on the Fisher Z transformed scores.) This result indicates that much more was learned from the Principled Revision than from the Original Text. That is, of the two recruit groups with equivalent abilities, the group who read the Principled Revision learned more than those who read the original. This result is consistent with the claim that we have developed a successful method for improving the learnability of Air Force instructional texts.

It is important to note that this result replicates the study reported as Experiment 2 of Britton and Gulgoz (1991). The corresponding correlation from that paper for the average correlation between the average experts' structure and the structure of the subjects who read the Original Text was .03, and the corresponding correlation for the Principled Revision was .21. These differed

by about as much in absolute terms as those for the present study, and the difference was reliable, $t(81) = 3.21$, p .002.

Performance on the test of inferences made while reading the test text also differed reliably between groups: it was 50% correct for recruits reading the Original Text and 57% for those reading the Principled Revision, $t(211) = 4.96$, $p<.001$. This replicates the results from Experiment 1 of Britton & Gulgoz (1991).

V. THE COMPUTER WRITING AIDS

This section describes our third broadly useful way of improving Air Force recruits' learning from Air Force instructional text. It describes two job aids to help Air Force writers improve Air Force instructional text. Texts improved with these aids will require much less of the individual difference abilities than the original versions of the texts, we believe. The result will be that texts revised with these aids will be learnable by recruits with lower levels of the abilities, as happened in the study reported in Section IV. The job aids are enclosed with the submitted version of this report. These are protytype versions; they run on IBMs or IBM clones. To increase their user-friendliness, further development work is needed.

Job Aid for Identifying Cognitive Gaps

Our first computer job aid calculates a measure of the number of cognitive gaps between successive sentences of any text. Its advantage over the Kintsch program used in Britton, Van Dusen, Glynn, and Hemphill (1990), and Britton & Gulgoz (1991) is that our job aid takes as input the raw text itself, rather than a propositionalized version of the text as is required by the Kintsch program. This is an important advantage because propositionalizing a text is so long and tedious a process that it is usually impractical except for research purposes. But our program will do the Kintsch-type analysis on the raw text itself; this

makes it feasible to analyze texts for cognitive gaps on a routine basis.

Why is the Program Useful?

It permits a quick and easy diagnosis of whether a text has a problem with cognitive gaps. We tested the program by analyzing the 8 military history texts used in Britton, Van Dusen, Glynn, and Hemphill (1990), and Graves et al. (1991). We then calculated the correlation between the recall scores and the number of gaps. Our writing aid predicted almost as well as the Kintsch program. The results are shown in Table 3 (compare to Table 1).

Table 3. Correlations between Cognitive Gaps (Counted by Job Aid) and Recall Learning

Study	Immediate Recall	Delayed 24 hours
Britton et al. (1990)	-.62	-.36
Graves et al. (1991)	-.42	-.53

In a second test of this job aid, we applied it to six other Air Force texts, to establish how widespread is the problem of cognitive gaps in Air Force texts. The texts were provided by the Air Force Extension Course Institute, and were on topics from courses in soldiering, welding, radio broadcasting, and vehicle maintenance. The results showed that cognitive gaps were found between more than half of the sentences and the previous text for all of the texts, as shown in the first column of Table 4. This shows that the problem of cognitive gaps is very common in Air Force texts. This means that the job aid would have wide application for revising Air Force instructional texts.

Table 4. Frequency of Cognitive Gaps in Six Air Force Instructional Texts
Cognitive gaps counted by Britton job aid

Text Topic	Cognitive Gaps (expressed as a percent of sentences)	Percent of Sentences using a different term for the same concept	Percent of Sentences with given -new problems
Field Fortifications	60	44	24
Metals	47	20	37
Brazing	50	8	8
Radio and Television Specialist	70	27	28
People and Electronic Journalism	58	18	11
Auto Frame Construction	60	18	13
Mean	58	21	21

Job Aid for Revising Text to Fill Cognitive Gaps

The second computer job aid was designed to help Air Force writers, educational specialists, and editors to revise Air Force instructional texts to fill the cognitive gaps.

Steps 1 and 2 of the program divide the text into numbered sentences, like the examples from the Air Force Vietnam text in Section II of this report. This is helpful for later analyses, because it is often useful for the reviser to refer to a sentence by its number.

Step 3 of the program prompts the user to divide compound sentences into two subsentences, and then the program labels those with numbers that reflect their sentence of origin. For example, if Sentence 3 is a compound sentence, the program will label its first part 3, and its second part 3A. This is useful because the separate parts of compound sentences are often separately subject to

different cognitive gaps.

The core of the job aid is steps 4-7. Steps 4-6 help the reviser find the terms that are subject to Corollary 1 of Principle 1 (described in Section IV). These are concepts for which the text uses several different terms. In our tests of the job aid, we have found that this problem is very common in Air Force texts, as shown in the second column of Table 4, being present in all of the texts, in about one-fifth of the sentences.

Step 7 (with Step 3) helps the reviser find the problems that can be repaired by Principle 2 (of Section IV) by identifying the "given" and "new" parts of each sentence. Then the program helps the revising process by printing out the given and new parts of each sentence in separate columns. In our test of the job aid, we found that given-new problems were common in Air Force texts, as shown in the third column of Table 4, being present in all of the texts, in about one-fifth of the sentences.

Future Prospects for These Job Aids

These job aids can be used by Air Force writers, educational specialists, and editors to improve Air Force instructional texts. The author would appreciate any suggestions for contacting units of the Air Force or other military services that might be interested in using the job aids.

VI. REFERENCES

Book Chapters

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), The psychology of learning and motivation (Vol. 8), 47-90.

Bollen, K. A. (1989). Structural equations with latent variables. New York: Wiley.

Britton, B. K., Gulgoz, S., & Glynn, S. M. (in press). Impact of good and poor writing on learners. To appear in B. K. Britton, A. Woodward, & M. Binkley (Eds.), Learning from textbooks. Hillsdale, NJ: Erlbaum.

Britton, B. K., Van Dusen, L., Glynn, S. M., & Hemphill, D. (1990). The impact of inferences on instructional text. In A. C. Graesser & G. H. Bower (Eds.), The psychology of learning and motivation (Vol. 25). New York: Academic Press.

James, L. R., Muliak, S. A., & Brett, J. (1982). Causal analysis: Assumptions, models, and data. Beverly Hills, CA: Sage.

Joreskog, K. G., & Sorbom, D. (1984). LISREL-V: Analysis of linear structural relations by method of maximum likelihood. Mooresville, IN: Scientific Software.

Kyllonen, P. C. (1991). CAM: A framework for cognitive ability measurement. In D. Detterman (Ed.), Current topics in human intelligence: Theories of intelligence. Morwood, NJ: Ablex.

Journal Articles

Ackerman, B. P. (1986). Referential and causal coherence in the story comprehension of children and adults. Journal of Experimental Child Psychology, 41, 336-366.

Bentler, P. M. (1990). Comparative fit indices in structural models. Psychological Bulletin, 107, 238-246.

Britton, B. K., & Gulgoz, S. (1991). Using Kintsch's computational model to improve instructional text: Effects of repairing inference calls on recall and cognitive structures. Journal of Educational Psychology, 83, 329-345.

Britton, B. K., Van Dusen, L., & Gulgoz, S. (1991). Reply to "A response to 'Instructional texts rewritten by five expert teams'". Journal of Educational Psychology, 83, 149-152.

Coleman, M., & Liau, T. L. (1975). A computer readability formula designed for machine scoring. Applied Psychology, 60, 283-284.

Graves, M. F., Prenn, M., Earle, J., Thompson, M., & Johnson, V. (1991). Commentary: Improving instructional text: some lessons learned. Reading Research Quarterly, 26, 110-122.

Haviland, S. E., & Clark, H. H. (1974). What's new? Acquiring new information as a process of comprehension. Journal of Verbal Learning and Verbal Behavior, 13, 512-521.

James, L. R., & Brett, J. M. (1984). Mediators, moderators, and tests for mediation. Journal of Applied Psychology, 69, 307-321.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. Psychological Review, 85, 363-394.

Lance, C. E. (1988). Residual centering, exploratory and confirmatory moderator analysis, and decomposition of effects in path models containing interactions. Applied Psychological Measurement, 12, 163-175.

Marsh, H. W., & Hocevar, D. (1988). A new, more powerful approach to multitrait-multimethod analysis: Application of second-order confirmatory analyses. Journal of Applied Psychology, 73, 107-117.

Miller, J. R., & Kintsch, W. (1980). Readability and recall of short prose passages: A theoretical analysis. Journal of Experimental Psychology: Human Learning and Memory, 6, 335-354.

Oakhill, J. (1986). Effects of time of day on the integration of information in text. British Journal of Psychology, 77, 481-488.

Schmidt, C. R., & Paris, S. G. (1983). Children's use of successive clues to generate and monitor inferences. Child Development, 54, 742-759.

DOD Reports

Ehrhart, R. C. (1978). History 202: Modern warfare and society. United States Air Force Academy.

U.S. Air Force Officers' Training Corps (1985). U.S. Airpower: Key to deterrence. Maxwell Air Force Base, Montgomery, AL: U.S. Air Force.

Unpublished Proceedings

Singer, M., Androsiak, P., Reisdorf, P., & Black, N. L. (1989). Individual differences in bridging inference processes. Paper presented at the Psychonomic Society Annual Meeting, Atlanta, GA.

Vosniadou-Papanicolas, S. (1982). Inferential processes in language comprehension: A psycholinguistic investigation of semantically positive and negative implicatory verbs. Unpublished dissertation, Clark University, Worcester, MA.

VII. APPENDIX

I. APPENDIX A: Original Air Force Instructional Text

(from Ehrhart, 1978)

Air War in the North, 1965

By the fall of 1964, Americans in both Saigon and Washington had begun to focus on Hanoi as the source of the continuing problem in the South. As frustration mounted over the inability of the ARVN to defeat the enemy in the field, pressure to strike directly at North Vietnam began to build. Although there was near unanimity among American officials over the aerial extension of the war into North Vietnam, serious differences arose over both the objective and the methods to be used.

Most members of the Johnson Administration believed bombing attacks would accomplish several things. They would demonstrate clearly and forcefully the United States' resolve to halt Communist aggression and to support a free Vietnam. At the same time, they would provide a boost for the sagging morale of the South Vietnamese. They would also make Hanoi pay an increasingly high price for supporting the Vietcong. Particularly among civilian advocates, the motivation for such a campaign was psychological rather than military, the primary objective not being Hanoi's capability but its willingness to support the war. "In a very real sense," explained Maxwell Taylor, "the objective of our air campaign is to change the will of the enemy leadership."

Those who stressed the psychological impact favored the strategy of "graduated response", a series of escalating attacks designed to show American resolve at relatively low cost while allowing the North Vietnamese leadership to realize the potential cost of supporting the war and thus cease that support before suffering the consequences. If they did not change their ways, severity

of the attacks would gradually be increased, with pauses between each increment to allow the enemy time to come to their senses. According to this theory, it would be the fear of future damage that would convince the enemy to cease their aggression. As late as the summer of 1965 Secretary of Defense Robert McNamara argued, "At any time, `pressure' on the Democratic Republic of Vietnam (DRV) depends not upon the current level of bombing but rather upon the credible threat of future destruction which can be avoided by agreeing to negotiate...." Graduated response assumed North Vietnam would not risk its fragile and limited industrial base in the face of overwhelming American power and would quickly succumb to the pressure thus exerted.

The Joint Chiefs of Staff (JCS), particularly Air Force Chief of Staff General John P. McConnell, adamantly opposed this approach. Graduated response, they contended, was a weak and indecisive strategy. It would not provide a clear symbol of our determination and resolve because of its incremental nature; nor would it damage Hanoi's war-making capability rapidly enough to be effective. Moreover, gradual escalation would give the enemy time to prepare psychologically and materially for each new step. The military leaders emphasized destruction of the enemy's capability to support the war rather than his will and favored brief but intensive operations to eliminate Hanoi's war-making capacity and to demonstrate the strength of the United States' commitment to win the war in a military sense.

In February 1965, President Johnson approved bombing operations against a selective list of targets in North Vietnam. While the government played down the significance of this step, Operation Rolling Thunder, which actually began in March, represented an important shift in the American approach to the war. The air raids following the Tonkin Gulf incident had been retaliatory strikes in response to specific attacks on American military forces. Rolling Thunder, in

contrast, was linked more broadly to the "larger pattern of aggression" by both the Vietcong and North Vietnam. Washington was now holding North Vietnam directly responsible for the war in the South.

From the beginning, Rolling Thunder was hedged with restrictions and limitations on sortie rates, targets, and even tactics and bomb loads. Rarely has the use of air power been so strictly controlled by civilians. Target selection, sortie rates, and routes were decided on a weekly (later bi-weekly) basis by a small group composed of the President, the Secretaries of Defense and State, and a handful of other senior officials only loosely guided by recommendations from CINCPAC and the JCS. Washington imposed restrictions of 30 and 15 nautical miles respectively around the capital city of Hanoi and the port of Haiphong. Intended to reduce the risk of superpower confrontation and to limit civilian casualties, these restrictions immediately limited the effectiveness of the campaign since the two cities were the center of North Vietnam's industrial base and the entry points for supplies from the Soviet Union and the People's Republic of China. Also off-limits was a strip thirty miles wide along the Chinese border.

By summer Rolling Thunder had grown in intensity. The possible target area had been extended from the 19th parallel to the 20th and then moved still further north. In addition to specified "hard targets" (stationary), strike forces had limited freedom to attack targets of opportunity in certain designated areas. Targets included military barracks, naval bases, railroad years, ferries, bridges, road repair equipment, and lines of communication (LOCs). The total number of sorties had risen to 900 a week, more than four times that of March.

Yet the bombing campaign had not achieved its objective. American policymakers had conceived of graduated response as a "low cost option with prospects for speedy, positive results." However, Hanoi showed no signs of

quitting. In the words of a CIA assessment of the early operations, "The strikes to date have not caused a change in the North Vietnamese policy....If anything, the strikes...have hardened their attitude."

The failure to obtain results led to a perceptible shift in the nature of the air war. While Rolling Thunder would continue, by July 1965 the objective slowly shifted from that of attacking psychological targets to the interdiction of North Vietnamese men and materials moving south. Whatever had been the hopes cherished by some officials in February and March, civilian leaders now conceived of the air campaign against the North as complementary to the ground war in the South. Reflecting this shifting objective, greater emphasis was placed on enemy lines of communication both in the panhandle of North Vietnam and along the Ho Chi Minh Trail in eastern Laos.

II. APPENDIX B: Principled Revision of Air Force Instructional Text

The 1965 Air War in North Vietnam

By the beginning of 1965, American officials in both South Vietnam and the U.S. had begun to focus on North Vietnam as the source of the continuing war in South Vietnam. The South Vietnamese army was losing the ground war against North Vietnam and this caused frustrations among the American officials. The frustrations led to pressure to bomb North Vietnam. The idea of bombing North Vietnam found support among nearly all the American officials. However, the civilian and military officials had serious differences over both the objective and the methods of the bombing attacks.

Most of both civilian and military members of the Johnson administration believed bombing attacks would accomplish several things. The bombing attacks would demonstrate clearly and forcefully the United States' resolve to halt communist North Vietnam's aggression and to support a free South Vietnam. At the same time, the bombing attacks would provide a boost for the South Vietnamese morale which was sagging because they were losing the war. The bombing attacks would also make North Vietnam pay an increasingly high price for supporting the war. Among the civilian officials, the motivation for the bombing attacks was psychological rather than military. For the civilian officials, the primary objective of the bombing was to break North Vietnam's willingness to support the war rather than its ability. Maxwell Taylor explained the civilian view: "The objective of our air campaign is to change the will of the enemy leadership."

In order to reach their psychological objective, the civilian officials favored the strategy of "graduated response." The graduated response strategy was designed to show at relatively low cost the American resolve to stop North Vietnamese aggression. Graduated response was a series of increasing attacks

with pauses after each increase to allow the North Vietnamese leaders to realize the potential cost of supporting the war. When they realized the potential cost, the North Vietnamese would cease supporting the war before suffering the consequences. If the North Vietnamese leaders did not cease supporting the war, the intensity of the bombing attacks would be further increased, with pauses between each attack to allow the North Vietnamese to realize the threat of future increases. According to the graduated response strategy, it would be the threat of future damage that would convince the North Vietnamese to cease their aggression. Secretary of Defense Robert McNamara explained that the pressure of the civilian strategy on North Vietnam "depends not upon the current level of bombing but rather upon the credible threat of future destruction which can be avoided by agreeing to negotiate...." The graduated response strategy assumed that North Vietnamese leaders would not risk the destruction of their fragile and limited industry by overwhelming American power. Therefore, they would quickly give in and cease their aggression.

The graduated response strategy did not find support among the military officials, particularly Air Force Chief of Staff General John P. McConnell. The military officials believed that the graduated response strategy was a weak and indecisive strategy. They believed the graduated response strategy would not be a clear symbol of U.S. determination and resolve because its intensity increased gradually. They also believed that this strategy would not damage North Vietnam's war-making ability rapidly enough to be effective. They believed that the pauses before each increase would give the North Vietnamese time to prepare psychologically and materially for the next bombing attack. Instead, the military officials thought it was more important to destroy North Vietnam's ability to support the war. In order to destroy North Vietnam's war-making ability and to show the strength of the U.S. resolve to win the war, the military

officials favored brief and intensive bombing attacks.

The graduated response strategy was President Johnson's choice. In February 1965, he approved bombing attacks against a selective list of targets in North Vietnam. The Johnson administration played down the importance of attacking North Vietnam directly. However, these bombing attacks, which began in March under the code name of Rolling Thunder, represented an important shift in the American approach to the war in Vietnam. Before Rolling Thunder, American attacks on North Vietnam, like the air raids following the Tonkin Gulf incident, had been retaliatory attacks in response to specific attacks on American military forces. In contrast to these retaliatory attacks, Rolling Thunder aimed at the "larger patter of aggression" by North Vietnam. By Operation Rolling Thunder, American officials were now holding North Vietnam directly responsible for the war in South Vietnam.

From the beginning, the bombing attacks of Operation Rolling Thunder had many restrictions and limitations. Restrictions included the rates of bombing attacks, targets, and even tactics and bomb loads. Aerial bombing attacks have rarely been so strictly controlled by civilians before. A small civilian group composed of the President, the Secretaries of Defense and State, and a handful of other senior officials made decisions about target selection, attack rates, and routes. The decisions of the civilian officials were only loosely guided by recommendations from military officials. The civilian officials imposed restrictions of 30 nautical miles aound the North Vietnamese capital city of Hanoi and 15 nautical miles around the port of Haiphong. Such restrictions were intended to limit civilian casualties and to reduce the risk of confrontation between the U.S. and the Soviet Union or China. However, these restrictions immediately limited the effectiveness of the bombing attacks because Hanoi and Haiphong were the center of North Vietnam's industry and these cities were also

the entry point for supplies from the Soviet Union and People's Republic of China. A restriction with a similar effect was a strip 30 miles wide along the Chinese border.

The bombing attacks of the graduated response strategy had increased in intensity by the summer of 1965. An example of the increase was that the possible target area had been extended from the 19th parallel to the 20th to cover a larger area in North Vietnam, and then the target area was moved still further North. Also, in addition to targets specified by the civilian officials, bombing forces had limited freedom to attack targets of opportunity in certain area. Such targets of opportunity included military barracks, naval bases, railroad yards, ferries, bridges, road repair equipment, and lines of communication (LOCs). The total number of attacks had risen to 900 a week, which was four times the number of attacks when Rolling Thunder began. Despite such an increase in the intensity of the attacks, the bombing attacks had not achieved the objective of breaking the will of the North Vietnamese. The civilian officials, who favored graduated response, had conceived of these bombing attacks as a "low cost option with prospects for speedy, positive results." However, the result of the attacks was that the North Vietnamese showed no signs of quitting. This was confirmed by a CIA report: "The strikes to date have not caused a change in the North Vietnamese policy....If anything, the strikes...have hardened their attitude."

The failure of the graduated response strategy to obtain positive results led to a clear shift in the nature of the air war. The shift was from the civilian strategy of attacking psychological targets to stopping forcefully the North Vietnamese soldiers and materials moving to South Vietnam. When Rolling Thunder began, before the civilian officials shifted to their new view, they had hoped that the graduated response strategy would stop the war. Since the war was

still going on, the civilian officials shifted to the view that the objective of the aerial bombing attacks against North Vietnam was to help win the ground war in South Vietnam. Reflecting this shift in the civilian view, greater emphasis was placed on bombing North Vietnam's lines of communication both in the panhandle of North Vietnam and along the Ho Chi Minh Trail in eastern Laos.